

Stylistique et ordinateurs P. Neruda et M. de Unamuno

Josse DE KOCK

La présente contribution est issue de la coordination d'une série d'initiatives individuelles et de la collaboration de plusieurs équipes de recherche (1). Les matériaux utilisés ont été rendus accessibles par ordinateur à l'Istituto di Linguistica Computazionale, à Pise (N. Catarsi, D. Ratti et T. Saba) sur l'initiative de R. Almela de l'Université de Murcie, A. Finzi de l'Université de Pise et R. Rosselli de l'Université de Florence, et au Centre d'Etudes Hispaniques et à la Section de Linguistique Appliquée de l'Université de Louvain (K.U. Leuven), par J. De Kock. La méthode d'analyse a été conçue, programmée et appliquée pour la première fois à Louvain par J. De Kock (2). Les données statistiques exploitées ci-dessous ont été produites au Laboratoire d'Analyse Statistique des Langues Anciennes de l'Université de Liège (L. Delatte, J. Denooz et S. Govaerts). Pour l'instant les résultats des calculs effectués sont disponibles sous la forme illustrée en annexe. Ces pages se veulent un hommage à la collaboration bénévole de chacun.

Cette contribution tend à illustrer une certaine recherche interdisciplinaire. D'essence littéraire, elle prône une approche linguistique, passe par la quantification et fait appel à l'ordinateur. En résumé elle repose, entre autres, sur les convictions méthodologiques suivantes (3).

En littérature, et *a fortiori* dans la poésie, l'étude des matériaux linguistiques est indispensable et prioritaire à tous les niveaux. Sans eux il n'y a pas d'oeuvre, de poème ni de message. La communication se différencie selon les matériaux

linguistiques utilisés, leur nature, leur fréquence, leur distribution et leur association entre eux et avec le monde référé.

L'analyse systématique, exhaustive et préalable de l'ensemble de l'oeuvre avalise l'interprétation de ses parties. Elle fournit les preuves concrètes et objectives de l'explication, alimente l'intuition détectrice et oriente les lectures postérieures. Elle est une garantie contre les interprétations déviantes que produisent facilement les mises en relation intentionnelles, et par là-même limitées, à partir de constatations occasionnelles et donc partielles.

La quantification, de pair avec l'automatisation qui la rend possible, permettent de brasser rapidement beaucoup de données à la fois et de façons variées. Elle permet de découvrir, de mettre en évidence et d'exprimer de manière précise et mesurée des relations qui restent d'habitude inconscientes au lecteur moyen ou perçues imparfaitement.

Grâce à l'ordinateur les données initiales et toutes leurs organisations intermédiaires sont toujours accessibles. L'automatisation bien comprise garantit à tout un chacun la possibilité de passer des résultats chiffrés au texte, de contrôler toutes les opérations effectuées, et de réutiliser les données enregistrées comme points de départ pour de nouvelles recherches.

A l'université, stylistique et explication littéraire n'ont de sens que si elles rendent compte de manière explicite et relative de tous les éléments qui composent le texte, et de leurs rapports. C'est ce travail de décomposition, d'isolations alternatives et de reconstitution qui les distingue de la lecture et de la synthèse impressionnistes - l'apanage du lecteur - qui les justifie, et qui les rend scientifiques.

Dans la pratique les caractéristiques d'une oeuvre littéraire se distinguent efficacement si on la compare de façon systématique, en détail et dans son ensemble, à d'autres productions linguistiques, sans qu'aucune d'entre elles ne soit, bien entendu, considérée comme la norme. Dans les pages suivantes nous résumons les étapes nécessaires d'une comparaison parmi d'autres sur un seul point et de façon simplifiée. D'autres aspects, d'autres comparaisons et d'autres procédures peuvent être envisagées.

Afin de réaliser la proposition qui précède, la formule suivante a été appliquée :

$$\frac{fT1 - fT2}{fT1 + fT2}$$

où f indique la fréquence relative d'une forme ou d'une classe de formes dans deux textes $T1$ et $T2$. Les résultats calculés oscillent entre 1 , 0 et -1 et sont donc comparables. La valeur calculée est positive ou négative selon que la forme examinée est plus ou moins fréquente dans $T1$ par rapport à $T2$. Une valeur élevée indique une différence de fréquence relative relativement grande ; une valeur proche de 0 qu'il n'y a pas ou peu de différence.

Pour contrôler la signification des résultats obtenus, l'écart réduit a été calculé parallèlement selon la formule de Reed :

$$\frac{fT1 - fT2}{\sqrt{\left(\frac{fT1 \cdot q1}{N1}\right) + \left(\frac{fT1 \cdot q2}{N2}\right)}}$$

où, en plus, N indique le nombre total d'occurrences dans chaque corpus et $q = 1 - f$. Les résultats peuvent être interprétés au moyen de tables. Rappelons que dans le cas d'une comparaison à deux variables, une valeur supérieure à $2,57$ révèle une différence significative avec seulement une marge d'erreur possible de 1% ; qu'en linguistique on admet souvent une marge d'erreur de 5% , soit le seuil de $1,97$; qu'au-dessous de ce seuil la valeur obtenue doit être considérée comme aléatoire.

Dans l'expérience commentée ici les deux formules ont été appliquées aux $1\ 000$ formes les plus fréquentes des quatre corpus suivants (4) :

- la poésie complète d'A. Machado (52 437 occurrences),
- la poésie complète de P. Neruda (354 784 occurrences),
- la poésie complète de P. Salinas (87 535 occurrences) et
- *Cancionero* de M. de Unamuno (76 779 occurrences),

comparés deux par deux dans cet ordre, et donnant lieu, par conséquent, à six séries de valeurs. Dans le cas présent les résultats sont présentés sous la forme de deux listes. Dans la première liste toutes les formes sont rangées selon la valeur de la différence de fréquence relative : ici de la valeur négative la plus élevée (-1) à 0 , et de 0 à la valeur positive la plus élevée ($+1$); la valeur de la différence de fréquence relative

est accompagnée de l'écart réduit. La deuxième liste reproduit les mêmes informations dans l'ordre alphabétique des formes examinées. Nous reproduisons en annexe quelques feuillets de ces deux listes. De gauche à droite l'on y trouve les comparaisons Machado/Neruda, Machado/Salinas, Machado/Unamuno, Neruda/Salinas, Neruda/Unamuno, et Salinas/Unamuno. Dans cette contribution nous nous bornerons à l'analyse sommaire de quelques formes de la poésie de Neruda et d'Unamuno.

Si on se laisse guider par la liste des formes ordonnées selon la valeur de la différence de fréquence relative les observations suivantes s'imposent, entre autres.

La forme à laquelle correspond la plus grande valeur de la différence de fréquence relative dans les six comparaisons est *Dios* : - 0,903 pour Neruda face à Unamuno. Cela veut dire que dans les poètes examinés aucune forme n'apparaît avec une fréquence aussi divergente que *Dios* dans Neruda et Unamuno. Vis-à-vis d'Unamuno l'absence ou l'extrême rareté du substantif ou de la notion *Dios* est un trait distinctif de la poésie de Neruda. Vis-à-vis de Neruda *Dios* est la forme la plus caractéristique de don Miguel.

Les formes de la même classe grammaticale de *Dios* qui suivent dans la liste selon l'ordre décroissant des valeurs de la différence de fréquence relative négatives jusqu'à - 0,500 (5) sont *cruz* (- 0,773), *lengua* (- 0,718), *eternidad* (- 0,705), *cara* (- 0,670), *España* (0,620), *abismo* (- 0,656), *nubes* (- 0,625), *pie* (- 0,589), *alma* (- 0,558), *palabra* (- 0,556), *rueda* (- 0,538), *historia* (- 0,534), *sol* (- 0,529), *verdad* (- 0,526), *padre* (- 0,520), *rey* (- 0,518), *risa* (- 0,514), et *universo* (- 0,513).

A l'inverse les substantifs plus fréquents dans Neruda que dans Unamuno, dans l'ordre décroissant des valeurs de la différence de fréquence relative, sont : *arena* (0,775), *ola* (0,712), *lluvia* (0,688), *viento* (0,671), *piedra* (0,662), *espuma* (0,652), *océano* (0,646), *nieve* (0,622), *bosque* (0,598), *manos* (0,587), *altura* (0,582), *patria* (0,575), *árbol* (0,571), *forma* (0,565), *selva* (0,545), *hombres* (0,521), *espacio* (0,513), *miel* (0,508) (6).

Pour trouver confirmation de la fréquence de ces substantifs il suffit de parcourir ne fût-ce que les titres et les premiers vers de Neruda comme de Unamuno. Voici un exemple pour chaque substantif cité dans l'ordre de leur valeur de la différence de

fréquence relative : Oda a la *arena* (NOE 1139) (7), Oda a la *ola* (TLO 1408), Oda a la *lluvia* (OE 1030), El *viento* es un caballo (VC 886), Al golpe de la ola contra la *pedra* indócil (CSA 1655), *Espuma* blanca, marzo en la Isla, veo (CC 1796), El gran *Océano* (CG 332), Primero resistió la tierra. / La *nieve* araucana quemó (CG 342), Oda al *bosque* de las Petras (TLO 1315), Oda a sus *manos* (NOE 1209), *Alturas* de Macchu Picchu (CG 312), *Patria*, te quieren repartir (CG 607), Aquí viene el *árbol*, el *árbol* (CG 353), No nadie reunirá tu firme *forma* (CC 1752), Entonces en la escala de la tierra he subido / entre la atroz maraña de las *selvas* perdidas (CG 316), Los *hombres* (CG 307), Mordido *espacio*, tropa restregada (R3 269), Espinas, vidrios rotos, enfermedades, llanto / asedian día y noche la *miel* de los felices (CSA 1679).

Dios de mi España, justicia (702) (8), Leño verde de la *cruz* (539), Mi íntimo abrigo, *lengua* castellana (363), ¿Es la *eternidad* que pasa (1175), Tu *cara*, mi cruz, me ampara, (546), España es un castillo (1519), Hubo tal soñador que en el *abismo* (1406), *Nubes* que pasais, mis cantos (114), Guíe a tu *pie* la cabeza, (66), La mente al *alma* convida (1216), La *palabra* me tortura (1019), Dale *rueda* al argandilla, (1049), *Historia*, colmo de histeria : (592), El *Sol* da de Dios ojeadas, (1379), ¿Qué es la *verdad* ? - A pregunta (187), *Padre* nuestro que estás, ¿como eres ? (116), Iba el *rey* pidiendo ranas (188), *Risa* que ríe y llora, (664), La ciencia cataloga al *universo* (1201).

Si l'on reconstitue les lemmes des formes dégagées par le calcul des valeurs de la différence de fréquence relative, en réunissant le singulier et le pluriel - quels qu'en soient la fréquence absolue ou le rang - aucune inversion de valeur ne se produit pour les lemmes par rapport à la forme déclinée originalement examinée. Si la forme déclinée la plus employée d'un substantif est relativement plus ou moins fréquente chez Neruda que dans Unamuno, le substantif en général y est aussi plus ou moins fréquent.

Mais à l'intérieur de la série positive comme de la série négative l'adjonction de la deuxième forme déclinée peut modifier la valeur de la différence de fréquence relative. En ajoutant les formes du pluriel au singulier de *árbol*, par exemple, la valeur de la différence de fréquence relative passe de 0,571 à 0,947 ; pour *pie* elle passe de - 0,589 à - 0,04 ; en additionnant la fréquence de *mano* à celle de *manos* elle passe de 0,587 à 0,25. Ces écarts sont parmi les plus élevés dans les deux listes.

Si la différence de fréquence relative pour *árbol* (0,571) passe à 0,947 pour la forme lemmatisée c'est parce qu'elle est égale à 1 pour *árboles* (101 occurrences pour Neruda et absent chez Unamuno), de - 0,589 pour *pie* à - 0,04 parce qu'elle est de 0,416 pour *pies* (258 fois dans Neruda et seulement 23 fois chez Unamuno), et de 0,587 pour *manos* à 0,25 à cause de - 0,072 pour *mano* (372 fois contre 93). Il est clair que les deux poètes n'utilisent pas également au singulier et au pluriel les substantifs relativement les plus fréquents chez chacun d'eux.

Lorsque, dans la série positive, la différence de fréquence relative des formes au pluriel est significative ($\sigma > 2,57$), la différence entre les valeurs de la différence de fréquence relative pour les formes du singulier et du pluriel d'un même lemme va dans le même sens : la valeur de la différence de fréquence relative est toujours plus éloignée de 0 pour le pluriel que pour le singulier. En d'autres mots, par rapport au singulier la forme du pluriel d'un même lemme est relativement plus fréquente chez Neruda que chez Unamuno. Il en est ainsi pour *arenas* (0,865 contre 0,775 pour *arena*), *lluvias* (1 contre 0,688), *piedras* (0,936 contre 0,662), *océanos* (1 contre 0,646), *bosques* (0,791 contre 0,598), *manos* (0,587 contre - 0,072), *alturas* (0,6 contre 0,582), *patrias* (1 contre 0,575), *árboles* (1 contre 0,571), *hombres* (0,521 contre 0,203), *espacios* (1 contre 0,513).

Dans la série négative la valeur de la différence de fréquence relative calculée pour le pluriel est rarement significative. Dans un cas, *pie(s)*, le pluriel est relativement plus fréquent que le singulier chez Neruda de façon significative (0,416 contre - 0,589) ; mais dans deux autres cas, *nube(s)* et *sol(es)*, c'est le contraire (- 0,625 contre - 0,348 et - 0,917 contre - 0,529) respectivement). L'on note cependant que pour sept lemmes [*dios(es)*, *cruz(ces)*, *lengua(s)*, *eternidad(es)*, *pie(s)*, *rueda(s)* et *padre(s)*] la valeur de la différence de fréquence relative est positive pour les formes du pluriel tandis qu'elle est négative pour le singulier.

Il existe, par conséquent, chez Neruda une nette tendance à user davantage du pluriel que du singulier des substantifs relativement plus fréquents dans sa poésie que chez Unamuno ; cette tendance se manifeste aussi, quoique dans une moindre mesure et non sans exceptions, pour les substantifs relativement moins fréquents chez Neruda que chez Unamuno.

Les tableaux 1 et 2 résument l'ensemble des données chiffrées pour le singulier, pour le pluriel ainsi que pour le lemme correspondant (9), des formes pour lesquelles une valeur de la différence de fréquence relative supérieure à 0,5, 1 et -1 exclues, a été calculée, dans l'ordre de grandeur de ces valeurs. La comparaison est faite du point de vue de Neruda.

Tableau 1

	s.	pl.	s + pl.	s.	pl.	s + pl.	Neruda		Unamuno	
Dios	-0,903	0,289	-0,840	17,39	1,91	16,83	79	60	338	7
cruz	-0,773	0,444	-0,692	8,30	1,81	7,82	55	24	93	2
lengua	-0,718	0,389	-0,640	7,07	1,47	6,65	56	21	74	2
eternidad	-0,705	1	-0,697	6,75		6,70	55	2	69	-
cara	-0,670	-0,213	-0,659	4,98	0,33	4,98	36	3	40	1
España	-0,620	-0,917	-0,680	8,32	2,13	8,59	102	1	112	5
abismo	-0,656	-0,396	-0,633	5,42	1,08	5,51	47	8	49	4
nubes	-0,348	-0,625	-0,540	9,03	5,09	5,40	38	49	17	53
pie	-0,528	0,416	-0,040	5,20	5,54	0,61	62	258	52	23
alma	-0,558	-0,644	-0,562	10,30	2,54	10,61	287	11	219	11
palabra	-0,556	-0,190	-0,404	6,51	2,07	6,42	116	151	48	48
rueda	-0,538	0,296	-0,355	3,77	1,42	2,95	43	34	31	4
historia	-0,534	-0,213	-0,512	4,77	0,57	4,77	70	9	50	3
sol	-0,529	-0,917	-0,542	10,26	3,01	10,66	333	2	234	10
verdad	-0,526	-0,072	-0,511	6,15	0,17	6,10	122	8	85	2
padre	-0,520	0,040	-0,480	5,17	0,13	5,00	89	15	61	3
rey	-0,518	-0,213	-0,484	3,61	0,57	3,62	44	9	30	3
risa	-0,514	-0,748	-0,542	3,53	2,06	4,03	43	4	29	6
universo	-0,513	-0,396	-0,510	3,65	0,54	3,69	46	2	31	1

Tableau 2

	s.	pl.	s + pl.	s.	pl.	s + pl.	Neruda		Unamuno	
arena	0,775	0,865	0,791	11,85	6,43	13,43	292	64	8	1
ola	0,712	-0,033	0,287	9,84	-0,38	4,93	248	186	9	43
lluvia	0,688	1	0,702	11,23		11,78	351	18	14	-
viento	0,671	0,039	0,591	14,11	0,13	11,27	588	15	25	3
piedra	0,662	0,936	0,746	13,38	15,01	18,97	546	278	24	2
espuma	0,692	0,146	0,574	8,53	0,67	7,78	225	31	10	5
océano	0,646	1	0,885	8,15		12,50	215	12	10	-
nieve	0,622	0,322	0,612	9,19	0,79	9,15	298	9	15	1
bosque	0,598	0,791	0,660	6,17	6,32	8,58	147	79	8	2
manos	-0,072	0,587	0,250	1,19	12,26	6,41	372	604	93	34
altura	0,582	0,600	0,586	5,84	3,11	6,62	140	37	8	2
patria	0,575	1	0,581	9,90		10,10	412	6	24	-
árbol	0,571	1	0,947	7,44		16,86	237	101	14	-
forma	0,565	0,346	0,534	5,51	1,24	5,53	133	19	8	2
selva	0,545	0,039	0,501	5,75	0,10	5,42	157	10	10	2
hombres	0,203	0,521	0,281	4,35	7,30	7,16	676	279	97	19
espacio	0,513	1	0,529	4,61		4,87	115	5	8	-
miel	0,508	-0,213	0,481	5,08	0,33	4,83	142	3	10	1

Les substantifs qui ont une valeur positive élevée chez Neruda (ici $> 0,5$), et qui sont donc caractéristiques de sa poésie, se réfèrent à la matière, les éléments et les forces de la nature : la terre et ses produits primaires (*arena, piedra, bosque, selva, árbol, miel*), l'eau sous ses diverses formes (*ola, lluvia, espuma, océano, nieve*), l'air (*viento*) (10), et à l'homme (*manos, hombres*). Les substantifs qui figurent dans la liste avec une valeur négative (ici $> - 0,5$), et qui sont donc caractéristiques de la poésie unamunienne, se réfèrent en majorité à des valeurs spirituelles ou à des notions abstraites : *Dios, cruz, eternidad, abismo, alma, verdad, Padre, universo ; lengua, palabra ; España, historia*. Neruda est le poète de la terre et des sens, Unamuno celui de l'esprit.

Le goût pour le concret de Neruda et le sens de l'abstrait d'Unamuno s'affirment aussi dans la tendance du premier à utiliser davantage que le second au pluriel les substantifs qui désignent des choses et même ceux qui expriment des notions, et, conjointement, l'inclination du second à employer davantage que le premier, au singulier les substantifs concrets, quelle que soit leur fréquence dans chaque poète : *alturas, árboles, arenas, bosques, cruces, dioses, espacios, eternidades, hombres, lenguas, lluvias, manos, océanos, padres, patrias, piedras, pies, ruedas*. En rendant les êtres, les éléments ou les objets et même les notions dénombrables Neruda met en évidence leur existence ; en les gardant au singulier Unamuno les convertit en symboles et vise davantage à leur essence.

Neruda : Volcanes, cicatrices,
 socavones,
 nieves ferruginosas,
 titánicas alturas
 desolladas,
 cabezas de los montes,
 pies del cielo,
 abismos del abismo,
 (...)

NOE 1154

 Porque a sudor y a fuego hicieron
 nacer los dioses de la piedra,
 (...)
 dioses pájaros, dioses serpientes,
 (...)
 América los levantó
 con mil pequeñas manos de oro,
 (...)

Por eso en las rocas crecieron
brazos y bocas, pies y manos,
la piedra se hizo monumento :
(...)

PCh 1709-1710

Unamuno :

Soñando cabecea, - soñando el lecho de la trilla,
y soñando la muela - que hace la harina,
y soñando la masa ya lluda - de pan de vida,
pan que hace mano, - y soñando la mano que echa semilla.
312

Guíe a tu pie la cabeza,
piensa bien a dónde vas;
y el corazón a tu mano,
de él tus obras sacarás.

66

Hablaba como un árbol ; en sus hojas
susurraba el aliento del Señor ;
endulzaba con su habla las congojas
en que el hombre madúrase el amor.

473

Mi bosque, de tu madera
un día han de hacer papel
de libros, nueva cantera
de otra torre de Babel;

1167,1-4

A celui qui entreprend l'étude d'une oeuvre la consultation des listes proposées permet de détecter instantanément quelques aspects distinctifs et de gagner ainsi un temps considérable. Au connaisseur elle permet de tester impressions et intuitions, en les confrontant avec des données établies de manière exhaustive et systématique, d'une grande précision et classées avec une totale objectivité. Les calculs proposés sont utiles à l'orientation des études littéraires comme à la vérification d'hypothèses de travail : ils alimentent et stimulent la recherche en suggérant un ordre d'exploration issu des données elles-mêmes ; ils fournissent des informations graduées de telle sorte que les résultats pour un ensemble donné sont immédiatement comparables à ceux obtenus de la même façon pour d'autres ensembles.

75

Ce qui précède n'est sans doute pas nouveau pour qui connaît Neruda et Unamuno. La démonstration n'en est pas pour autant inutile à nos yeux. Dans l'étude des textes littéraires, et dans les sciences humaines en général, il ne s'agit pas tellement de découvrir de nouvelles choses insoupçonnées, que de fonder la perception du texte – la connaissance – sur des données objectives, de telle sorte qu'elle devienne explicable et transmissible.

La coïncidence entre la démonstration et le savoir tout court devrait inciter à un peu plus de foi dans la quantification, trop souvent et hâtivement jugée inapte pour traiter des oeuvres de l'esprit humain. Une certaine quantification peut rendre compte de la production linguistique, et donc littéraire. Elle est particulièrement utile pour les réalisations qui se font entièrement ou en partie de façon inconsciente, ou échappent à des degrés variables aux capacités de la mémoire humaine par leur ampleur dans l'espace et dans le temps. Pour l'ensemble de la production linguistique elle est une voie de connaissance susceptible de doubler la connaissance intuitive ou raisonnée. Il va de soi que la quantification ne dispense personne du retour au texte, et que tout texte appelle une interprétation.

Notes

- (1) Il a été rendu compte pour la première fois de ce projet au *I Congreso internacional sobre el español de América*, 4-9/10/1982, à Puerto Rico, sous le titre "*Análisis lingüístico de textos literarios*", à paraître dans les Actes.
- (2) J. De Kock, *Elementos para una estilística computacional*, en colaboración con D. Geens, N. Delbecque, C. Gómez Molina y R. De Keyser, Madrid, 1983, 2 t.
- (3) Pour plus de détails et pour d'autres applications voir Josse De Kock, *Elementos para una estilística computacional*, o.c.
- (4) Le choix des textes découle en partie du souci de varier les expériences (antérieurement *Cancionero* de Unamuno avait été confronté aux articles de journal du même auteur et à la liste de fréquence de A. Juillard) et en partie du hasard : Machado, Neruda et Salinas sont trois poètes disponibles en machine à Pise.
- (5) Les seuils d'analyse fixés dans cette contribution : 1 000 formes dans chaque oeuvre, valeurs de la différence de fréquence supérieures à 0,500, formes substantives, etc., n'ont pas de justification méthodologique ou technique particulière. Il n'y a pas d'inconvénient à examiner plus de 1.000 formes et des formes avec une différence de fréquence relative inférieure à 0,5 mais significatives (> 1,97), ou d'autres formes de la même catégorie ou d'autres catégories.

- (6) Il convient de rappeler que les formes à la fois caractéristiques de la poésie de Neruda et d'Unamuno vis-à-vis de Machado et de Salinas, ou de l'un d'eux, (également fréquentes dans Machado et Salinas) ne figurent pas dans cette liste. Pour les dépister il suffirait d'examiner aussi les autres comparaisons. Pour caractériser une oeuvre complètement il y aurait lieu de mener à terme autant de confrontations qu'il y a de textes à lui opposer.
- (7) Les abréviations sont celles de *Obras completas*, Buenos Aires, 1962, p. 1851.
- (8) Le chiffre réfère au numéro du poème dans *Cancionero*.
- (9) Sur le plan méthodologique la différence de fréquence parfois considérable et relativement systématique des formes du singulier et du pluriel des substantifs examinés invite à remettre en question la lemmatisation globale et préalable telle qu'elle est pratiquée habituellement dans les études littéraires. La lexicographie et la lexicologie, du côté des linguistes, et la traduction ou la documentation automatiques du côté des informaticiens, ont intoxiqué à ce point l'informatique des textes qu'une lemmatisation paraît à la plupart nécessaire et primordiale. Et de fait, point d'index tenu en estime qui n'ait sacrifié à la lemmatisation, qu'elle ait été opérée artisanalement ou automatiquement, intégralement ou en partie. A ce niveau peu d'interventions sont pourtant aussi coûteuses, en temps et en argent. Le prix en semble d'autant plus exorbitant qu'elle s'avère souvent peu utile, voire contre-indiquée. La lemmatisation entraîne une perte d'information considérable, utile si pas essentielle pour le stylisticien comme pour le linguiste. La perte enregistrée ci-dessus à propos des substantifs (singulier et pluriel) est minime, comparée à la perte que la lemmatisation des verbes, par exemple, provoque.
- (10) *Fuego*, qui évoque le quatrième élément, a également une valeur positive : 0,345.

Annexe 1

-1.0000		vuelto			vuelta	
	0	- 3.16	0	0	- 5.09	0
-1.0000		vuelven			vueltas	
	0	- 3.31	0	0	- 3.00	0
-1.0000		última			vuestro	
	0	- 4.58	0	0	- 3.60	0
-1.0000		últimas			yedra	
	0	- 3.31	0	0	- 3.60	0
-1.0000		último			yerba	
	0	- 4.69	0	0	- 4.35	0
-1.0000		única			zamora	
	0	- 3.31	0	0	- 3.46	0
-1.0000		0	0	0	última	
	0				- 4.69	0
-1.0000		0	0	0	único	
	0				- 3.60	0
-0.9037					dios	
					- 17.39	
-0.8827						dios
						- 16.82
-0.8694						ay
						- 12.81
-0.8351		sí				
		- 12.68				
-0.8264		ahora				
		- 9.33				
-0.8162		eso				
		- 8.27				
-0.8149			ay			
			- 11.88			
-0.8103	ahora					
	- 14.56					
-0.8066		cuerpo				
		- 10.28				
-0.7730					cruz	
					- 8.30	
-0.7518				sí		
				- 12.61		
-0.7478					san	
					- 6.19	
-0.7449						mas
						- 8.12
-0.7436				tan		
				- 15.45		
-0.7329						soñar
						- 6.83
-0.7315					ay	
					- 11.79	
-0.7256		arriba				
		- 5.93				
-0.7234			queda			
			- 6.45			
-0.7232	tal					
	- 10.66					
-0.7198		fin				
		- 6.72				
-0.7187				carne		
				- 7.99		
-0.7185					lengua	
					- 7.07	

-0.7179		estás		
-0.7177		- 5.78		le
-0.7167			fin	- 14.13
-0.7162			- 6.31	
-0.7156		ti	eterno	
-0.7149		- 12.19	- 5.89	
-0.7076		mí		pobre
-0.7057		- 10.21		- 7.59
-0.6986	pronto			eternidad
-0.6963	- 9.06			- 6.75
+0.7054				hizo
+0.7060			fuego	- 5.34
+0.7106		dios	+ 14.12	igual
+0.7116		+ 7.30		+ 6.14
+0.7127		hermanos		
+0.7151		+ 4.48		ola
+0.7155				+ 9.84
+0.7265				nadie
+0.7289				+ 10.97
+0.7336				forma
+0.7381		clara		+ 5.67
+0.7392		+ 6.26		detrás
+0.7557		jardín		+ 6.57
+0.7568		+ 5.09		
+0.7584	dios	tarde		
+0.7634	+ 7.87	+ 9.44		
+0.7685	claro	fría		
+0.7719	+ 6.72	+ 5.18		
+0.7752		poeta		
+0.7789		+ 6.04		gran
+0.7852				+ 10.54
				nosotros
				+ 6.65
				entonces
				+ 13.71
				arena
				+ 11.85
				cuerpo
				+ 10.24
			tal	
			+ 13.24	

+0.7959		campos + 6.53				
+0.8014	vieja + 6.78					
+0.8106		viejo + 6.89				
+0.8279					hacia + 8.99	
+0.8376				allí + 17.57		
+0.8503		hacia + 7.80				
+0.8513					allí + 10.49	
+0.8531				hacia + 18.06		
+0.8700	campo + 9.26					
+0.8976		campo + 9.32				
+1.0000	águila + 3.00	águila + 3.00	álamos + 4.69	ésta + 7.00	árboles + 10.05	árboles + 4.00
+1.0000	álamos + 4.69	álamos + 4.69	árboles + 2.64	éste + 7.00	ésta + 7.00	abajo + 7.07
+1.0000	abejar + 2.44	abejar + 2.44	abejar + 2.44	abeja + 6.92	éste + 7.00	abiertas + 3.00
+1.0000	abel + 3.00	abejas + 3.46	abejas + 3.46	abejas + 6.78	abajo + 9.00	abierta + 3.74
+1.0000	abrí + 2.44	abel + 3.00	abel + 3.00	abismo + 6.85	abeja + 6.92	abiertos + 3.60
+1.0000	abril + 6.56	abrí + 2.44	abierta + 2.64	abrió + 7.00	abejas + 6.78	abren + 3.16
+1.0000	acacias + 2.64	acacias + 2.64	abrí + 2.44	aceite + 7.55	abierta + 7.07	abril + 5.83
+1.0000	acaso + 4.00	adolescente + 2.44	abril + 6.56	adelante + 6.08	abrió + 7.00	abrir + 3.74
+1.0000	adolescente + 2.44	adónde + 2.44	acacias + 2.64	agonía + 7.07	aceite + 7.55	acaban + 3.46

